DOCUMENT RESUME

ED 096 374 UD 014 543

AUTHOR Jaeger, Richard M.; Freijo, Tom D.
TITLE Race and Sex as Concomitants of Teachers' Accuracy in Evaluative Rating of Students.
SPONS AGENCY University of South Florida, Tampa.
PUB DATE Apr 74
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Academic Achievement; Compensatory Education Programs; Elementary Schools; *National Surveys; Perception; *Racial Differences; Sex Differences; Student Characteristics; *Student Evaluation; Student Needs; Student Teacher Relationship; Teacher Attitudes; *Teacher Characteristics

ABSTRACT

This study examines the hypothesis that the accuracy of teachers' evaluations of their students is higher if (1) the teacher and the student are of the same race, and (2) the teacher and the student are of the same sex. Data were collected in the 1969 Survey on Compensatory Education, a survey of national scope sponsored by the U.S. Office of Education. The survey secured data from school superintendents, elementary school principals, and elementary school teachers. Teachers answered questions about themselves and their classes in responding to a "Teacher Questionnaire," and answered questions about individual pupils in their classes in responding to several "Pupil Questionnaires." Survey data were secured from a nationally-representative sample of public school systems and elementary schools which provided services, supported in whole or in part through Title I of the Elementary and Secondary Education Act of 1965, during the 1968-69 school year. Only students and teachers from grade 4 were considered in the present study. The results of this study suggest that much of the literature on the topic of race and social distance may not be applicable to classroom teachers. "White female teachers rating blacks" were most accurate, and "black teachers rating blacks" were least accurate. The most accurate evaluations were made by female teachers and by whites rating other whites. (Author/JM)

ED 096374

RACE AND SEX AS CONCOMITANTS OF TEACHERS' ACCURACY IN EVALUATIVE RATING OF STUDENTS [1]

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Richard M. Jaeger

Tom D. Freijo

University of South Florida

INTRODUCTION

The Research Hypothesis

It has been well documented (McDonald, 1965; Davis, 1964) that accurate evaluation of students' strengths and weaknesses is fundamental to effective teaching. Lacking accurate evaluation, a teacher cannot determine the effectiveness of past instruction, nor can (s)he determine needs for remediation.

While teachers gather some evaluative information through objective devices such as standardized achievement tests, a great deal of evaluation depends upon less formal data-collection procedures (Thorndike and Hagen, 1969). Teachers often make prescriptive decisions for students on the basis of unquantified observations and ratings, for example. It is the latter kinds of evaluations that this study concerns.

This study examines the hypothesis that the accuracy of teachers' evaluations of thier students is higher if

(1) the teacher and the student are of the same race,

and

(2) the teacher and the student are of the same sex.

UD 0[illegible]

The term "accuracy of evaluation" is defined quite specifically. The definition can be motivated by describing a hypothetical situation. Suppose that a large group of teachers of both sexes and different races provides ratings of students' progress on several school skills and social behaviors. The rated group of students is also large, and consists of both sexes and several races. Suppose further that the school skills and social behaviors on which students are rated, while not completely uncorrelated, show considerable independence of development in the population.

An "accurate evaluation" is then characterized as one in which a teacher considers each rated behavior independently of every other rated behavior, and an "inaccurate evaluation" as one in which a student is rated holistically, without regard to her/his differences across rated behaviors. This kind of inaccuracy in rating was recognized as a serious problem by Willingham and Jones (1958), and labeled "composite halo".

Given the situation described above, the following compound hypothesis is tested: The accuracy of teachers' evaluations of students' progress depends upon the race and sex of the teacher and the race and sex of the student. More specifically, the following ordering for teachers' evaluations of students' progress is hypothesized (from most accurate to least accurate):

(1) Teachers evaluating students of the same race,

(2) black female teachers evaluating white students,

(3) black male teachers evaluating white students,

(4) white male teachers evaluating black students,

(5) white female teachers evaluating black students.

This hypothesized ordering is consistent with a large body of sociological literature, much of which is reviewed below.

One can generalize further from previous research, and hypothesize an

ordering on accuracy of rating for all combinations of sex and race of teacher and sex and race of student. The additional hypotheses are based on extrapolations of specific research findings; some refutations of these hypotheses could therefore be expected. The complete ordering is as follows (from most accurate ratings to least accurate):

(1) Teachers evaluating students of the same race and sex,

(2) Teachers evaluating students of the same race

(3) White teachers evaluating white students

(4) Black teachers evaluating black students

(5) Teachers evaluating students of the same race but opposite sex

(6) Black female teachers evaluating white students

(7) Black male teachers evaluating white male students

(8) Black male teachers evaluating white female students

(9) White male teachers evaluating black students

(10) White female teachers evaluating black students.

Supporting Research

While there appears to be no previous research on the relationship between classroom teachers' evaluation accuracy and the sex and race of teachers and students, there is a considerable body of literature for more general populations that relates the similarity between evaluators and those being evaluated, and evaluation accuracy. The literature appears to support two contentions. First, evaluative raters show greater interest toward ratees similar to themsleves (Kagan, 1967); interpret their perceptions of ratees in terms of their self-perceptions (Stagner, 1948); are more sensitive to those characteristics of ratees which conform to their self-perceptions (Fernsterheim and Tresselt, 1953); and better recall those characteristics of ratees which conform to thier own personalities (McLaughlin, 1970).

Second, accuracy of rating is positively related to the similarity of the rater and the ratee; the more similar the two, the more accurate the rating. McLaughlin (1970) found that ratees were given higher ratings by similar raters than by dissimilar raters. Christensen (1970) found that ratees were given more accurate ratings by similar raters than by dissimilar raters.

The research reviewed above suggests that all teachers, regardless of sex or race, might be expected to rate students of their own sex and race more accurately than they do those of different sex and/or race. In addition, one can hypothesize a rank ordering on teachers' accuracy of student evaluation, which depends on the sex and race of teachers in relation to those of the students being rated. Several studies which base their findings on the concept of social distance support the hypothesized ranking given above.

Bogardus developed a scale in 1926 for assessing the relative perceived social distance between groups. The scale has been used by many investigators to study the relative perceived distance between groups of differing nationality and race. One of the most consistent findings of these studies is the large perceived distance between whites and blacks in the United States (Kock, 1946; Bogardus, 1958; Meltzer, 1941; Ames, 1968; Hines, 1968). Several studies considered the relationship between sex and social distance. Bogardus (1969), for example, found that women perceived themselves to be considerably more distant from those of different race and nationality than did men. Ames (1968) and Landis (1966) confirmed Bogardus findings. Research supports the contention that the perceived social distance of whites from blacks is larger than the perceived social distance of blacks from whites. Kock (1946) found that reciprocal feelings of distance from children of the other race existed among both white students and black students throughout their school years. She found further that the perceived distance of white students from blacks was larger

than the perceived distance of black students from whites. Bogardus (1958) found the same relationships to exist among black and white college students and college graduates, between the ages of 18 and 35. These studies were consistent in their finding that whites place blacks at the extreme end of the social distance scale, while blacks view themselves as less socially distant from whites. The reported findings represent averages, of course, and individual differences exist within each racial group. Hines (1968) found that blacks rated various national-origin and racial groups in the following order on a social distance scale (from closest to farthest): Anglos, Mexican-Americans, American-Indians. He also found that among blacks over 21, women rated themselves as significantly closer to whites than did men.

These results support the hypothesis that blacks, because of their perceived social "closeness" to whites, will rate whites more accurately than whites will rate blacks. Christensen (1970) presents further evidence to substantiate this hypothesis. He compared blacks' ratings of whites and whites' ratings of blacks with corresponding self-ratings on a number of personality traits. Christensen found that black raters were just as accurate as white raters in rating whites, but white raters were far less accurate than black raters in rating blacks.

Significance of the Research

According to McDonald (1965) teaching is a decision-making process in which the essentials of the educative act consist of (1) formulation of the goals of the learning experience, (2) development of a plan for the instructional strategy and (3) formulation of a plan for evaluating the effect of the strategy. Similarly, Newell, Shaw and Simon (1958) present a decision-making model which can be applied to teaching. Their model highlights the

fundamental importance of accurate evaluation in the formulation of educational goals and the development of instructional strategies. McDonald stresses the importance of accuracy in teachers' perceptions of students' strengths and weaknesses:

> "Such inaccuracies in our perception of pupils will probably affect their development adversely because of the way we will treat them. Inaccurate perceptions will also interfere with planning of appropriate learning experiences. We might be handicapped in our understanding of the goals likely to motivate a particular student; we might err also in evaluating the factors likely to stimulate pupil change or in our estimates of the factors inhibiting change." McDonald, p. 532.

Ojemann and Wilkinson (1939) reported the effects on student growth of providing teachers with additional, accurate information about students. At the end of a one-year experimental study, the investigators found that students of teachers given additional information had significantly higher adjustment, than did students of teachers with less information. In a similar but more recent study, Hoyt (1955) found that providing teachers with additional information on the characteristics of their students had no effect on student achievement, but did result in positive changes in student attitudes. These studies show that the amount of accurate information a teacher has on (her) his students does affect the students in important ways. As noted by Tagiuri and Petrullo (1957), the teacher's perception of students' characteristics is important information in defining the educational problem to be solved.

It would seem therefore, that determining the extent to which teachers of one race or sex fail to make accurate judgments of the educational progress

of students of another race or sex is a significant subject of inquiry. It would fail to be so, only if the racial and sexual isolation of teachers and students in our schools was nearly complete, and was likely to continue to be so.

PROCEDURES

Data Source

Data used to investigate the hypothesis that the accuracy of teachers' evaluations of student progress depends upon the race and sex of the teacher and the race and sex of the student were collected in the 1969 Survey on Compensatory Education; a survey of national scope sponsored by the U.S. Office of Education.

The 1969 Survey on Compensatory Education (hereafter called the Survey) secured data from school superintendents, elementary school principals, and elementary school teachers. Teachers answered questions about themselves and their classes in responding to a "Teacher Questionnaire", and answered questions about individual pupils in their classes in repsonding to several "Pupil Questionnaires". Survey data were secured from a nationally-representative sample of public school systems and elementary schools which provided services, supported in whole or in part through Title I of the Elementary and Secondary Education Act of 1965, during the 1968-69 school year. In addition to a Pupil Questionnaire and a Teacher Questionnaire, the Survey used a Principal Questionnaire and a School District Questionnaire. In total, 172 multi-part questions were asked on these questionnaires; 71 on the Pupil Questionnaire, 38 on the Teacher Questionnaire, 44 on the Principal Questionnaire and 19 on the School District Questionnaire. All of the data secured through the four questionnaires can be linked, in the sense that responses

for a student can be tied uniquely to responses for her/his teacher, her/his principal and her/his school district.

The Instrument

Five questions from the 1969 Survey on Compensatory Education--two from the Teacher Questionnaire and three from the Pupil Questionnaire--were used in conducting the research reported here. Questions used are as follows:

Question 2 on the Teacher Questionnaire:

What is your sex?

#Male

#Female

Question 6 on the Teacher Questionnaire:

Are you a member of one of the national minority groups (Racial, or national origin groups which are a minority of the national population.) listed below?

#Yes

#No

If yes, please indicate which one:

#American Indian

#Negro

#Oriental

#Spanish-surnamed American (Persons of Cuban Descent, Mexican Descent, Puerto Rican Descent, Spanish Descent)

Question 2 on the Pupil Questionnaire:

What is this pupil's sex?

#Male

#Female

Question 10 on the Pupil Questionnaire:

Is this pupil a member of any of the following national minority groups? (Racial or national origin groups which are a minority of the national population.)

#Yes

#No

If yes, which one?

#American Indian

#Negro

#Oriental

#Spanish-surnamed American (Persons of Cuban Descent, Mexican Descent, Puerto Rican Descent, Spanish Descent)

Question 41 of the Pupil Questionnaire:

Please indicate the change in this pupil's academic performance and behavior since you first became his teacher during the 1968-69 school year. Rate this pupil on each item listed [below], taking into consideration how he performed when you first became his teacher this school year and how he performs now. (Assume that the school year started in the Fall of 1968 and does not include a summer session.)

(Note: On the questionnaire, the item is arranged in matrix form. There are six possible options for each of twenty-one performances and behaviors. The options, and the performances and behaviors are listed below.)

Item options:

Large change for the better

Some change for the better

- # No change (change desirable)
- # No change (change not necessary)
- # Some change for the worse
- # Behavior not observed

Pupil performance and behaviors:

- # Care in handling school property
- # Responsibility in completing class assignments
- # Attentiveness in class
- # Creativity
- # Relationships with adults
- # Relationships with other pupils
- # Amount of disruptive behavior
- # Understanding oral instructions
- # Accuracy in self evaluation
- # Self concept
- # Dress habits
- # Anxiety
- # Attendance
- # Reading proficiency
- # Math proficiency
- # Oral expression
- # Awareness of current affairs
- # Educational aspirations
- # Liking for his teacher
- # Independent learning
- # Understanding written instructions

Questions 2 and 6 on the Teacher Questionnaire and questions 2 and 10 on the Pupil Questionnaire were used as classification variables. That is,

they were used to establish linked files of students and teachers of both sexes and two racial groups. A Likert scale was imposed on question 41 on the Pupil Questionnaire, in order to quantify teachers' responses. "Large change for the better" was scaled as "5", "Some change for the better" was scaled as "4", "No change (change desirable)" was scaled as "3", "No change (change not necessary)" was scaled as "2", "Some change for the worse" was scaled as "1", and "Behavior not observed" ratings were treated as missing data.

The Sample

The Survey Sample. Since the data that were available for the present research are prescribed by the design of the 1969 Survey sample, its structure will be described first. The primary sampling units used in the Survey were school districts with enrollments exceeding 299 students that received funds under Title I, ESEA during the 1968-69 school year. A sample of 438 school districts was selected from a universe of 9236, using systematic random sampling within each of four enrollment strata. Sizes of district samples within the enrollment strata were determined using Neyman (1938) optimal allocation. Enrollment boundaries, population sizes and sample sizes for the four strata were as follows:

Stratum	Enrollment within School District	Number of Title I Districts in Stratum	Sample Size
1	40,000 or more	92	91
2	9,000 to 39,999	658	124
3	3,000 to 8,999	1917	121
4	300 to 2,999	6569	102
	TOTALS	9236	438

Sampling of schools within selected districts was, with slight modification, accomplished using a systematic random procedure with a sampling fraction of 1:1.4. The sampling frame used within each selected district was a list of all schools with at least one of grades 2, 4, or 6 that provided services supported in whole or in part under Title I, ESEA.

Within sampled schools, all principals and all teachers in grades 2, 4 and 6 were sent survey questionnaires, provided at least 15 pupils were enrolled in a sampled grade. Sampled teachers were asked to complete questionnaires for themselves and for a sample of 3 to 6 students in their classes. Each sampled teacher was provided with a precise systematic sampling procedure for the selection of students within her/his class.

This sampling procedure resulted in the selection of 2,920 schools, 22,067 teachers, and 104,036 students in grades 2, 4 and 6. The Survey response rates exceeded ninety percent.

The Research Sample. While the entire sample of teachers and students available from the 1969 Survey could have been used in the study, to do so would have been unnecessarily wasteful. There was no desire in the present research to estimate totals for the national population, as was true in the 1969 Survey. Only students and teachers from grade 4 were considered in the present study.

The procedure used for sampling the available data was based upon two factors. First, the essential randomness of the data file was to be preserved. Since the data were arranged by schools within districts, districts within strata and strata within states, a systematic sampling procedure was used to assure proportional representation across these classification variables, while preserving randomness. Second, the size of the Survey data file afforded the opportunity to investigate the stability of the analytic findings using jackknife procedures as suggested by Miller (1968).

Prior to sampling, the classification variables described above (questions 2 and 6 on the Teacher Questionniare and 2 and 10 on the Pupil Questionnaire) were used to create separate but linked files of teacher data and student data, classified by race and sex; sixteen files were thus created. Since it was found that only twenty-three white male students were rated by black male teachers, and only fourteen white female students were rated by black male teachers, these files were eliminated from further analyses. The remaining fourteen files were then joined to compose groups of data, corresponding to eight of the ten hypotheses listed above (two hypotheses could not be investigated, since data for black male teachers rating white students were not available in sufficient quantity). The groups of data were then sampled systematically, with sampling fractions chosen so as to provide samples of approximately 1000 ratings of students by teachers. In two groups no additional sampling was performed since the numbers of ratings available were either less than or close to 1000. The eight groups, the numbers of cases in each group, and the sampling fractions used are shown in Table 1.

Insert Table 1 about here

Because single classification variables were used to designate some groups and multiple classification variables were used to designate others, the groups listed in Table 1 are not mutually exclusive. For example, the group "Teachers rating students of the same race and sex" is a subset of the group "Teachers rating students of the same race". For purposes of analysis, it was necessary that the samples used be non-overlapping. Therefore, starting-points for the systematic sampling procedures were chosen so as to provide mutually exclusive samples.

Data Analysis

Heuristic Discussion. The purpose of the procedures used for analysis of data was to examine the relative structural complexity of teachers' ratings of students, for white and black teachers of both sexes, and white and black students of both sexes. Put another way, it was assumed that variables such as "Care in handling school property", "Accuracy in self evaluation", and "Reading proficiency", although positively correlated across students, would exhibit some unique variation. It was hypothesized, for example, that a white teacher rating a white student would be better able to perceive the unique character of these variables than would a white teacher rating a black student. If hypotheses such as these were correct, the result would be a more complex relational structure among the rated variables, when the race or sex of teacher and student were the same, than when the race or sex of teacher and student were different.

These hypotheses were examined using two different analytic procedures: Principal components analysis (Kaiser, 1958) and non-metric multidimensional scaling (Shepard, 1962; Kruskal, 1964a). Each procedure was applied independently to teachers' ratings of students, when teachers and students were grouped by sex and race, as listed in Table 1. The details of these procedures are discussed below.

Principal Components Analysis. Principal components analyses of teachers' ratings of students were completed independently, using the sampled data for each of the groups listed in Table 1. Intercorrelations among the twenty-one rated student behaviors were computed first, and the resulting correlation matrix was then analyzed through the principal components algorithm provided by the BMD03M computer program (Dixon, 1965).

The criterion of rating accuracy used in this study was, as noted above, highly specialized. In particular, teachers' ratings were judged to be more accurate, the lower the composite halo effect exhibited. For the principal components analyses, degree of composite halo effect was operationally defined by the proportion of standardized variance among the twenty-one dimensions rated, that was accounted for by the first principal component. Thus the larger the variance attributable to a single factor, the larger the composite halo effect.

For purposes of comparing the samples of ratings of each of the groups listed in Table 1 then, it would have been sufficient to compute the proportion of variance accounted for by the first principal component of their respective correlation matrices. A comparison among populations represented by these samples (rather than the samples themselves) were of primary interest, however. The analyses were therefore more complex. It was desired to formally test hypotheses on the pairwise equality of compound halo effect for the populations represented by each pair of groups listed in Table 1. Since the sampling distribution of the proportion of variance accounted for by a single principal component is not known, approximation techniques were required. The procedure employed was the jackknife technique (Miller, 1968). By dividing sample data into overlapping subsets and computing estimates for each, the jackknife procedure provides an estimate of a population parameter of interest, and in additon, provides an estimate of the variance of the estimator. These two estimates -- the parameter estimate and its variance estimate -- can be combined through standard statistical procedures to formally test hypotheses.

Once jackknife estimates of the proportion of variance accounted for by a single principal component and estimates of the variances of the jackknife

estimates were computed for each group listed in Table 1, pairwise comparisons were made using Bonferroni t-statistics (Miller, 1966). The twenty-eight pairwise comparisons were made with an overall experimental error rate of five percent.

Non-metric multidimensional scaling. The relative composite halo effect among ratings by the eight groups listed in Table 1 was also investigated using multidimensional scaling. The advantage of multidimensional scaling is its weak assumption on the measurement properties of teachers' ratings. While the principal components procedure assumes an interval level of measurement for teachers' ratings of students, the multidimensional scaling procedure requires only ordinal measurement.

The operational definition of degree of composite halo effect used with multidimensional scaling was the size of the "stress" value resulting from an attempt to fit the relationships among the twenty-one rated variables into a unidimensional space. Stress is a measure of badness of fit, as defined by the Kruskal multidimensional scaling algorithm (1964a). That is, the higher the stress value, the worse the fit of the data to a single dimension (and concomitantly, the smaller the composite halo effect).

The procedures used with multidimensional scaling were nearly identical to those described above for the principal components analyses. As in the principal components analyses, the intercorrelations among the twenty-one rated student behaviors were first computed separately for each of the eight groups listed in Table 1. These correlations were treated as similarities among the twenty one variables, and were used as inputs to Kruskal's MDSCAL IV computer program (1964b).

Since the sampling distribution of the stress statistic is unknown, the

jackknife procedure was once again employed to test pairwise hypotheses of identical composite halo effect among the eight groups listed in Table 1. Again, Bonferroni t-statistics were used to maintain an overall experimental error rate of five percent.

RESULTS

Two independent statistical techniques were used to test hypotheses in this study. Since the two techniques provided differing results, the findings for each procedure will be presented separately.

Factor Analysis

As defined in this study, an inaccurate evaluation is one which exhibits a large amount of composite halo effect. For factor analyses, the degree of composite halo effect has been operationally defined as the proportion of standardized variance among rated behaviors accounted for by the first principal component. An accurate evaluation is therefore characterized by a factor analysis in which a relatively small proportion of variance is accounted for by the first principal component; an inaccurate evaluation is characterized by a factor analysis in which a relatively large proportion of variance is accounted for by the first principal component.

Using these definitions, the groups of teachers and students corresponding to the hypotheses of this study are listed in decreasing order of accuracy of evaluative ratings in Table 2. For comparative purposes, the hypothesized ordering is also listed. The hypothesized ordering of groups and the ordering

---

Insert Table 2 about here.

---

found through factor analysis differ considerably. To be consistent with the research cited above, the "White females rating blacks" group should have produced the least accurate evaluations. Conversely, the "Blacks rating blacks" group should have produced the most accurate evaluations. In fact, the evaluations provided by these groups were directly opposite to the rankings hypothesized; "White females rating blacks" were most accurate, and "Blacks rating Blacks" were least accurate. In terms of individual characteristics, the most accurate evaluations were made by female teachers and by whites rating other whites. Since a large percentage of the white teachers were female, the "Whites rating whites" group reflects a predominance of females that may outweigh the racial classification.

The eight groups listed in Table 1 give rise to 28 pairwise comparisons of evaluation accuracy. When these comparisons were made, only two pairs of groups were found to differ significantly in the proportions of standardized variance accounted for by the first principal component. The two significant differences (at the 0.05 level) were found between the "White females rating blacks" group vs. the "Blacks rating blacks" group and between the "Whites rating whites" group vs. the "Blacks rating blacks" group. For the "Blacks rating blacks" group, the first principal component accounted for a larger percentage of variance in both comparisons, indicating theirs was the less accurate evaluation. The statistics associated with all 28 comparisons are listed in Table 3.

Insert Table 3 about here

Multidimensional Scaling

In a multidimensional scaling analysis, a high degree of composite halo

effect (inaccurate evaluation) is operationally defined by a relatively low value of stress when the relationship among behavioral ratings are mapped into a single dimension. Therefore, the lower the stress value associated with a one-dimensional solution, the less accurate the evaluation; conversely, the higher the stress value, the more accurate the evaluation.

Using this definition, the groups of teachers and students corresponding to the hypotheses of this study are listed in decreasing order of accuracy of evaluative ratings in Table 4. Once again, the hypothesized ordering is listed for comparative purposes. The hypothesized ordering of groups and the ordering found through multidimensional scaling are somewhat different. Consistent with the factor analysis results cited above, the "White females rating blacks" group prodcued the most accurate evaluations. The "Blacks rating blacks" group, which the theory predicted to be most accurate, was among the least accurate.

Insert Table 4 about here

When multidimensional scaling results were used to make pairwise comparisons of evaluation accuracy for the eight groups listed in Table 1, significant differences were found for only two pairs of groups. The "Whites rating whites" group was found to be significantly less accurate in its evaluations (at the 0.05 level) than the "White males rating blacks" group. The "Whites rating whites" group was also found to provide significantly less accurate evaluations than did the "White females rating blacks" group. Statistical results for all 28 pairwise comparisons are shown in Table 5.

Insert Table 5 about here

CONCLUSIONS

Factor Analysis

1. Overview

The results of this study do not agree with the results predicted by the cited sociological literature. However, the role of research is not just to substantiate hypothesized relationships, but to identify unexpected findings and reconcile these with established knowledge. To that end, a rationale that explains the results of this study is developed in this section. The rationale is based upon three main contentions: (1) that teachers are different from the population on which most of the perceived social distance research is based, (2) that male teachers are different from female teachers in ways that would affect their evaluation accuracy, and (3) that black teachers are different from white teachers in ways that would affect their evaluation accuracy.

These three main contentions are elaborated and supported with references to previous studies. An explanation of the ordering of dyads obtained in this study, consistent with research on the sociology of teachers and teaching, is then presented.

2. Contentions

1) In reviewing the literature dealing with social distance, the relative difference between the perceived distances of whites from blacks and blacks from whites was well established; however, the social distance studies reviewed did not restrict thier populations to teachers. There is evidence to substantiate the contention that teachers may be different from the general population (and even other college graduates) in ways which would affect their perceived social distances, because of two factors -- (1) selection and (2) training. Several studies have indicated that those who select teaching as a career posess a system of values which places a high priority on the social

worth of others (Mason, et. al, 1959; Brembeck, 1962; Rosenberg, et.al, 1957; Gottlieb, 1961; Spaeth, 1959). In thier training of teachers, many colleges have as a high priority to sensitize prospective teachers to the worth of individuals and the acceptance of individual differences (Spindler, 1968). Therefore,because of self-selection and specialized training, it may be assumed that teachers exhibit different patterns of perceived social distance than the populace as a whole. If this were the case, the ralationships between sex, race and perceived social distance hypothesized for the general population would not apply to the population of teachers.

2) Another explanation for the discrepency between the hypothesized and actual orderings of groups on accuracy of evaluation might be the presence of one or more confounding variables which are differentially represented in the population of teachers, when categorized by race and sex. One such variable is socio-economic status (SES). A review of literature showed that more than one sociological study supports the contention that SES is more important than race in determining perceived social distance; e.g., white middle-class persons have less perceived social distance from middle-class blacks than from lower-class whites.(Gordon, 1964; Landis, 1966).

3) Other potentially confounding variables are the levels of training and/or motivation of teachers represented in the different groups. For example, it is well documented that many males enter classroom teaching with expectations of attaining positions in educational administration while a very high percentage of females have no aspirations to move from classroom teaching to some "higher" position (Mason, et. al, 1959; Colombotos, 1962; Graebell and Olson, 1973).

4) More than one sociologist has noted a phenomenon exhibited by blacks who reach middle-class status. It has been asserted that because of the oppresion and negative prejudice to which lower-class blacks are subjected,

blacks who reach middle-class status often disassociate themselves from lower-class blacks. Moreover, an expression of their disassociation is a level of criticism of lower-class blacks that exceeds the criticism exhibited by middle-class whites. Hentoff (1965) discusses this phenomenon and cites several supporting sources.

5) It has been proposed by several educators and sociologists that for many decades blacks have been "trapped" in a self-perpetuating and inferior educational system. It is claimed that black teachers have been provided an inferior education, teach predominantly black students, and that some of these students are trained to become teachers by black colleges, thus perpetuating the system. This proposition was substantiated by the survey and case studies mandated by the Civil Rights Act of 1964 (Coleman, et. al, 1966; USOE, 1966). If black teachers do receive inferior training, they would be expected to be less accurate in their evaluation of students.

3. Explanation of Findings

If the contentions described and supported above are assumed to be true, they may provide a basis for explaining the ordering of groups on accuracy of evaluation as reported for the factor analysis procedure.

Why would white female teachers evaluating blacks provide the most accurate evaluations? Teaching has long been a respectable terminal career for middle-class females. Teachers in this group rarely use their teaching position as a stepping stone to a "higher" position in the educational hierarchy. It is reasonable to assume then, that middle-class teachers are, as a group, interested in classroom teaching and are, for the most part, highly motivated. Their training, too, is probably specific to the classroom teacher position, in contrast to administratively-oriented males who may seek specific training

in administration. If the ability to evaluate effectively is one of the skills that defines good classroom teaching, it is reasonable to conclude that persons specifically trained for that role position, interested in the position, and highly motivated toward effective functioning in the position, will evaluate more effectively than those who are not. Why then do black teachers exhibit such a high degree of composite halo when rating black students (Group 6), even though ninety percent of the black teachers are female? There are two possible reasons: (1) The training received by black female teachers is probably inferior to the training received by white female teachers in the majority of cases, and the evaluation accuracy of black female teachers may be reduced accordingly; (2) At the time the data reported here were collected, most black female teachers were teaching classes that were predominantly or exclusively black. A large proportion of low socio-economic status students would be expected in these classes. It has been contended that blacks, upon reaching middle-class status, often perceive themselves to be socially distant from lower-class blacks. It is reasonable to assume that this phenomenon would be exhibited by middle-class black female teachers. A combination of inferior training and large perceived social distance, then, could explain the relative inaccuracy of black teachers in rating black students.

Black female teachers were relatively accurate in their evaluations of white students (Group 2). This result is not consistent with the general sociological theory discussed above, and requires further explanation. Again, two possibilities are suggested by the sociology of the teaching profession: (1) The number of black female teachers who teach white students is far smaller than the number who teach black students. It is not unreasonable to assume that black females who teach white students do so in predominantly white schools, and are placed in those schools because they are good teachers (Sexton, 1964).

(2) Black female teachers may feel no need to alienate themselves from lower-class white students, although, as has been contended above, they do feel the need to alienate themselves from lower-class black students. It is also reasonable to assume that the socio-economic status of student bodies in predominantly white schools is higher than that in predominantly black schools; the perceived social distance of black female teachers from middle-class white students would be smaller than their perceived social distance from lower-class blacks.

The "Whites rating whites" group (Group 5) is ranked third in order of accuracy of evaluation. This group contains male as well as female teachers, whereas the groups ranked first and second in evaluation accuracy contained only female teachers. It has been claimed that white male teachers are relatively disinterested in teaching, are relatively poorly motivated, and often enter the teaching profession with the expectation of moving on to administrative positions. Why, then, does the "Whites rating whites" group rank third in evaluation accuracy. The composition of the group probably explains its high ranking, since more than ninety percent of the white teachers are female. The small proportion of male teachers in the group would only slightly diminish the evaluation accuracy of a group primarily composed of white female teachers.

Group 1 (Teachers rating students of the same race) ranked fourth in accuracy of evaluation. This intermediate position might properly reflect a combination of elements that would enhance evaluation accuracy (the superior training, interest, and motivation of white female teachers) and elements that would diminish evaluation accuracy (inferior training of black teachers, relative lack of motivation and disinterest of male teachers).

The "White males rating blacks" group (Group 3) was also intermediate in evaluation accuracy; its rank was five. According to the contentions cited

above, the group is composed of relatively disinterested teachers who are evaluating students from whom they feel less socially distant than do middle-class black female teachers. These offsetting factors would suggest an intermediate position on accuracy of evaluation.

Sociological arguments explicit to the classroom teacher role do not suggest the relative ordering observed for groups 3, 7 and 8 ("White males rating blacks", "Teachers rating students of the same race and sex", and "Teachers rating students of the same race but opposite sex", respectively). However, on the operational variable (proportion of variance accounted for by a single factor) used to define evaluation accuracy, these groups were separated by only eight-hundredths of one percent -- a statistically insignificant separation.

Multidimensional Scaling

The arguments used to explain the results on accuracy of evaluation found through factor analysis can, with two exceptions, be used to explain the results found through multidimensional scaling. The "Black females rating whites" group was found to be second most accurate in evaluation using the factor analysis procedure, but was found to be fifth most accurate using the multidimensional scaling procedure. The results of the multidimensional scaling analysis for this group are generally consistent with the original hypotheses of this study, but do not conform to the explanation given above for the factor analysis results. The "Whites rating whites" group was found to be third most accurate in evaluation using the factor analysis procedure, but was found to be the least accurate group using the multidimensional scaling procedure. This result is not consistent with the original hypotheses of this study, nor can it be explained by the teacher-specific sociology used to

explain the factor analysis results. The negative value of "stress" obtained for this group defies statistical explanation, although it is an allowable consequence of the jackknife procedure. The best estimate of "stress" for the "Whites rating whites" group is therefore in doubt.

Comparison of Factor Analysis Results and Multidimensional Scaling Results

The most consistent result of the two methods of analysis was the number one ranking on accuracy of evaluation of the "White females rating black students" group. Considering all of the groups, however, the two methods of analysis produced marked inconsistencies. The Spearman rank-order correlation between the two sets of rankings was .38. The inordinately low ranking of the "Whites rating white students" group in the multidimensional scaling analysis (a possible statistical artifact) materially reduced the rank-order correlation between the two sets of rankings. Had this group been ranked identically on accuracy of evaluation for both methods of analysis, the rank-order correlation between groups would have been .69.

Because the results obtained for the "Whites rating white students" group with the multidimensional scaling analysis defy explanation, and because the results obtained with factor analysis are supported by sociological theory, it is suggested that greater confidence be placed in the factor analysis results.

CONCLUDING REMARKS

The results of this study suggest that much of the literature on the topic of race and social distance may not be applicable to classroom teachers. This preliminary finding could be confirmed through a direct investigation of the social distance perceptions of classroom teachers. A study of the

elements of teacher training curricula that affect social distance perceptions would also be worthwhile.

It has also been suggested that social class may be a more important factor than race in determining social distance perceptions. While this suggestion derives more from the post hoc interpretation of the findings of this study than from the data explored, it is consistent with a body of sociological literature. A study of the relative contributions of race and social class to the evaluation accuracy of classroom teachers is now in progress.

REFERENCES

Ames, Richard G.; Moriwaki, Sharon Y.; and Basu, A.K., "Sex Differences in Social Distance: A research Report," Sociology and Social Research, 52, 1968, pp. 280-289.

Bogardus, Emory S., "Racial Distance Changes in the U.S. During the Past Thirty Years," Sociology and Social Research, 43, 1958, pp. 127-135.

Bogardus, Emory S., "Race Reactions by Sexes," Sociology and Social Research, 43, 1959, pp. 439-441.

Brembeck, Cole S., The Discovery of Teaching, Englewood Cliffs, New Jersey: Prentice-Hall Inc. 1962, pp. 38-43

Christensen, Larry, "Person Perception Accuracy as a Function of Ethnic Group and Familiarity," Perceptual and Motor Skills, 31, 1970, pp. 510.

Coleman, James S, et. al, Equality of Educational Opportunity, U.S. Department of Health, Education, and Welfare, Office of Education, Washington, D.C.: Government Printing Office, 1966.

Colombotos, John L., Sources of Professionalism: A Study of High School Teachers, United States Office of Educational Cooperative Research Project No. 330, Ann Arbor: Department of Sociology, University of Michigan, 1962, pp. 414, 415, 417.

Davis, F.B., Educational Measurements and Their Interpretation. Belmont, California: Wadsworth, 1964.

Dixon, W.J., ed., Biomedical Computer Programs, Los Angeles: Health Services Computing Facility, School of Medicine, University of California, 1965.

Fensterheim, H. and Tresselt, M.E., "The Influence of Value Systems on the Perception of People," Journal of Abnormal and Social Psychology, 48, 1953, pp. 93-98.

Landis, Judson, R.; Datwyler, Darryl; and Dorn, Dean S., "Race and Social Class as Determinants of Social Distance," Sociology and Social Research, 51, 1966, pp.78-86.

Mason, S. Ward; Dressel, Robert J.; and Bain, Robert K., "Sex Role and the Career Orientations of Beginning Teachers," Harvard Educational Review, 29, 1959, No. 4.

McDonald, F., Educational Psychology, 2nd Ed., Belmont, California: Wadsworth Publishing Company, 1965.

McLaughlin, Barry, "Similarity, Recall, and Appraisal of Others," Journal of Personality, 38, 1970, pp. 106-110.

Meltzer, H., "The Development of Children's Nationality Preferences, Concepts, and Attitudes," Journal of Psychology, 11, 1941, pp. 343-358.

Miller, Rupert G., Simultaneous Statistical Inference, New York: McGraw-Hill, 1966.

Miller, Rupert G., "A Trusty Jackknife," Annuals of Mathematical Statistics, 35, 1968, pp. 1594-1605.

Newell, A.; Shaw, J.C.; Simon, H.A., "Elements of a Theory of Human Problem Solving," Psychological Review, 65, 1958, pp. 151-166.

Neyman, J., "On the Two Different Aspects of the Representative Method; the Method of Stratified Sampling and the Method of Purposive Selection," Journal of the Royal Statistical Society, 97, 1934, pp. 558-606.

Ojemann, R.H.; and Wilkinson, F.R., "The Effect on Pupil Growth of an Increase in Teachers' Understanding of Pupil Behavior", Journal of Experimental Education, 8, 1939, pp. 143-147.

Rosenberg; Suchman; and Goldsen, Occupations and Values, New York: The Free Press, 1957.

Sexton, Patricia, Education and Income, New York: The Viking Press, Inc. 1964.

Gordon, Milton M., Assimilation in American Life, New York: Oxford University Press, 1964.

Gottlieb, David, "American Graduate Students -- Some Characteristics of Teachers and Researchers," Journal of Educational Psychology, 52, 1961, No. 5.

Graebell, Leon C.; and Olson, John A., "Role Dissatisfaction and Career Contingencies Among Female Elementary Teachers," Journal of the Student Personnel Association for Teacher Education, 11, 1973, No. 4.

Hentoff, Nat, The New Equality, New York: The Viking Press, 1965.

Hines, Ralph H., "Social Distance Components in Integration Attitudes of Negro College Students," The Journal of Negro Education, 37, 1968, pp.23-30.

Hoyt, K.B., "A Study of the Effects of Teacher Knowledge of Characteristics on Pupil Achievement and Attitudes Toward Class Work," Journal of Educational Psychology, 46, 1955, pp. 302-310.

Izard, C.E., "Personality Similarity, Positive Affect, and Interpersonal Attraction," Journal of Abnormal and Social Psychology, 61, 1960, pp. 484-485 (b).

Kagan, J., "On the Need for Relativism," American Psychologist, 22, 1967, pp. 131-142.

Kaiser, H., "The Varimax Criterion for Analytic Rotation in Factor Analysis," Psychometrika, 23, 1958, pp. 187-200.

Koch, Helen L., "The Social Distance Between Certain Racial, Nationality, and Skin-Pigmentation Groups in Selected Populations of American School Children," Journal of Genetic Psychology, 68, 1946, pp. 63-95.

Kruskal, J.B., "Multidimensional Scaling by Optimizing Goodness of Fit to Non-metric Hypotheses," Psychometrika, 29, 1964a, pp. 1-27.

Kruskal, J.B., "Nonmetric Multidimensional Scaling: A Numerical Method," Psychometrika, 29, 1964b, pp. 219-246.

Shepard, R.N., "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function," Psychometrika, 27, 1962, Part I, pp. 125-140, Part II, pp. 219-245.

Spaeth, Joe L., "Values Orientations and Career Preferences of Graduate Students," (Paper read at the Chicago meetings of the American Sociological Association, 1959).

Spindler, George Dearborn, "The Acculturation of the School Teacher," in Chilcott, et. al eds. Readings in the Socio-Cultural Foundations of Education, Belmont, California: Wadsworth Publishing Company, Inc., 1968.

Stagner, R., "Psychological Aspects of Industrial Conflict: I. Perception," Personnel Psychology, 1, 1948, pp. 131-144.

Tagiuri, R. & Petrullo, L., eds., Person Perception and Interpersonal Behavior, Stanford, California: Stanford University Press, 1957, pp. 277-288.

Thorndike, Robert and Hagen, Elizabeth, Measurement and Evaluation in Psychology and Education, 3rd ed. New York: John Wiley and Sons, 1969.

U.S. Department of Health, Education, and Welfare, Office of Education. Equality of Educational Opportunity. Washington, D.C.: Government Printing Office, 1966.

Willingham, W.W. and Jones, M.B., "On the Identification of Halo Through Analysis of Variance," Educational and Psychological Measurement, 18, 1958, pp. 403-407.

FOOTNOTES

1. This work was supported, in part, by a grant from the University of South Florida Division of Sponsored Research. We wish to express our appreciation of Mr. Ronald Register who helped in analyzing the data.

TABLE 1

GROUPING OF TEACHERS AND PUPILS, INITIAL SAMPLE SIZES, AND SAMPLING FRACTIONS

Group	Number of Cases Available from Survey	Sampling Fraction
Teachers rating students of the same race	23,262	1/20
Black female teachers rating white students	662	all cases used
White male teachers rating black students	1,375	all cases used
White female teachers rating black students	4,275	1/4
White teachers rating white students	14,069	1/10
Black teachers rating black students	9,193	1/10
Teachers rating students of the same race and sex	11,346	1/10
Teachers rating students of the same race but opposite sex	11,916	1/10

TABLE 2

HYPOTHESIZED AND ACTUAL LISTING OF GROUPS IN DECREASING ORDER OF ACCURACY OF EVALUATION. ACTUAL ORDERING BASED ON FACTOR ANALYSIS

Group No.	Hypothesized Order	Group No.	Actual Order
7	Teachers rating students of the same race and sex	4	White female teachers rating black students
1	Teachers rating students of the same race	2	Black female teachers rating white students
5	White teachers rating white students	5	White teachers rating white students
6	Black teachers rating black students	1	Teachers rating students of the same race
8	Teachers rating students of the same race but opposite sex	3	White male teachers rating black students
2	Black female teachers rating white students	7	Teachers rating students of the same race and sex
3	White male teachers rating black students	8	Teachers rating students of the same race but opposite sex
4	White female teachers rating black students	6	Black teachers rating black students

TABLE 3

PROPORTIONS OF VARIANCE ACCOUNTED FOR BY FIRST PRINCIPAL COMPONENT IN VARIOUS TEACHER-STUDENT GROUPINGS, AND PAIRWISE TESTS OF BETWEEN-GROUP DIFFERENCES

Group No.	Group	Proportion of Variance Accounted for by First Principal Component	Proportion of Variance Accounted for by First Principal Component (Jackknife Estimate)	Standard Error of Jackknife Estimate
1	Teachers rating students of the same race	.395	.3923	.0141
2	Black female teachers rating white students	.389	.3863	.0173
3	White male teachers rating black students	.417	.4178	.0129
4	White female teachers rating black students	.375	.3720	.0112
5	White teachers rating white students	.383	.3866	.0125
6	Black teachers rating black students	.434	.4367	.0065
7	Teachers rating students of the same race & sex	.420	.4182	.0122
8	Teachers rating students of the same race but opposite sex	.415	.4186	.0141

Groups	t-statistic	Groups	t-statistic	Groups	t-statistic	Groups	t-statistic
1-2	.269	2-3	-1.461	3-5	1.826	4-8	-2.581
1-3	-1.357	2-4	.694	3-6	-1.307	5-6	-3.877*
1-4	1.108	2-5	.146	3-7	- .023	5-7	-1.913
1-5	.316	2-6	-2.712	3-8	.042	5-8	-1.753
1-6	-2.828	2-7	-1.515	4-5	- .922	6-7	1.342
1-7	-1.388	2-8	-1.434	4-6	-4.992*	6-8	1.164
1-8	-1.315	3-4	2.676	4-7	-2.791	7-8	.021

* significant at .05 level; critical value = 3.82

TABLE 4

HYPOTHESIZED AND ACTUAL LISTING OF GROUPS IN DESCENDING ORDER OF ACCURACY OF EVALUATION, ORDERING BASED ON MULTIDIMENSIONAL SCALING

Group No.	Hypothesized Order	Group No.	Actual Order
7	Teachers rating students of the same race and sex	4	White female teachers rating black students
1	Teachers ratings students of the same race	1	Teachers rating students of the same race
5	White teachers rating white students	3	White male teachers rating black students
6	Black teachers rating black students	8	Teachers rating students of the same race but opposite sex
8	Teachers rating students of the same race but opposite sex	2	Black female teachers rating white students
2	Black female teachers ratings white students	7	Teachers rating students of the same race and sex
3	White male teachers rating black students	6	Black teachers rating black students
4	White female teachers rating black students	5	White teachers rating white students

TABLE 5

STRESS VALUES FOR ONE-DIMENSIONAL STUDENT RATINGS BY TEACHERS IN VARIOUS TEACHER-STUDENT GROUPINGS, AND PAIRWISE TESTS OF BETWEEN-GROUP DIFFERENCES

Group No.	Group	Stress	Jackknife Estimate of Stress	Standard Error of Jackknife Estimate
1	Teachers rating students of the same race	.443	.4097	.0738
2	Black female teachers rating white students	.399	.2964	.0455
3	White male teachers rating black students	.431	.4094	.0336
4	White female teachers rating black students	.441	.4167	.0387
5	White teachers rating white students	.280	-.0629[1]	.1053
6	Black teachers rating black students	.410	.2804	.0908
7	Teachers rating students of the same race & sex	.382	.2956	.0406
8	Teachers rating students of the same race but opposite sex	.357	.3255	.0268

Groups	t-statistic	Groups	t-statistic	Groups	t-statistic	Groups	t-statistic
1-2	1.306	2-3	-1.998	3-5	4.273*	4-8	1.938
1-3	.004	2-4	-2.014	3-6	.301	5-6	-2.584
1-4	.084	2-5	3.131	3-7	2.159	5-7	-3.175
1-5	3.674	2-6	- .158	3-8	1.953	5-8	-3.574
1-6	1.105	2-7	.013	4-5	4.275*	6-7	.153
1-7	1.354	2-8	.551	4-6	1.381	6-8	.476
1-8	1.072	3-4	- .143	4-7	2.174	7-8	.614

[1] This value was treated as 0.0 in the calculation of statistics

* Significant at the .05 level; critical value = 3.82